PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	How weight change is modelled in population studies can affect research findings: empirical results from a large-scale cohort study
AUTHORS	Paige, Ellie; Korda, Rosemary; Banks, Emily; Rodgers, Bryan

VERSION 1 - REVIEW

REVIEWER	Anna Peeters Baker IDI Heart and Diabetes Institute Australia
REVIEW RETURNED	16-Mar-2014

GENERAL COMMENTS	This article addresses a topic of importance – is there an optimal way to analyse weight change, and what are the potential implications of using different measures. I think the authors have selected their outcomes well but would suggest inclusion of another option. Major suggestions: There is a body of literature suggesting that in analyses of change in a variable, that variable should not be adjusted for at baseline (see for example Shea American Journal of Epidemiology 1998), known as the horse racing effect. However, the role of baseline weight can be important. So, I would recommend you re-run your continuous analyses also predicting follow up weight and adjusting for baseline weight. This may be relevant for the apparently discordant results for absolute versus relative continuous change for women versus men who have very different absolute weight and weight change. Your findings may also reflect the non-linear association between some predictors and weight change, such that loss is not the same as gain. Another option to consider including would be looking only at weight gain, by excluding major weight losers from the analysis. This may be an option to retain the power of a continuous analysis but not be confounded by predictors of weight loss. With your large power you could also consider a sensitivity analysis restricting the analysis to never smokers as they are a substantial confounder of the relationship which can never be fully accounted for through adjustment by simple smoking status.

REVIEWER	David A. Shoham Dept. of Public Health Sciences Loyola University Chicago Maywood, IL, USA
REVIEW RETURNED	02-Apr-2014

GENERAL COMMENTS	1. This paper is a well-written useful addition to the literature, as it compares different cutpoints for defining stable vs.

	weight change by education level. I do not know of other studies that have compared different cutpoints, although other studies have compared absolute vs. relative change. 2. Although noted in the limitations, the use of self-reported weight is a major concern, because we might expect the variance to be correlated with education level (ie, the more educated may more accurately report weight and thus weight change) and also to be non-differentially biased (for example, if the less education underreport by a greater amount than the higher-educated). I would therefore caution against over-interpretation of the results as favoring cutpoints over continuous methods. As the authors point out, we would expect a systematic under-reporting to wash out in the analyses, and while I agree, correlation of variance with education may be a greater concern, as results may be driven as much by random noise. In other words, if there is more variance in measured among the less educated than among the highly educated, then we will see greater numbers or less educated crossing upper- and lower- thresholds for defining weight change, just due to chance. This will cancel out in a linear regression, but emerge in a categorical regression. 3. Tables 5 and 6 seem unnecessary. We don't need to know such detail about the relationship between potential confounders and weight change. On the other hand, presenting adjusted, partially adjusted, and fully adjusted models are helpful. Unless the authors can clarify the Hausman test for me, this might require someone trained in econometrics and more familiar with these tests. This might be a perfectly appropriate application of the Hausman method, I am just not certain. I am a bit confused about the use of the Hausman test here and the only reference provided for the test is the Stata users manual and I have not seen it used in epidemiologic research before. My understanding of this test is that it tests consistency and efficiency, where a more efficient estimator is compared to a (presumed) consistent estimator. In cases where I have seen the test applied, it is used to test whether additional variance parameters need to be included in the model (for example, random effect vs. fixed effect), a robust estimator is needed (eg, due to outliers), or irrelevant alternatives in multinomial models are independent of one another (the IIA assumption). The test is really one of whether a model that is more efficient (but has unknown consistency), can substitute for a known-to-be-consistent (but less efficient) model. The main issue they are examining in this paper, however, is whether two sets of parameter estimates are statistically different, when they would be presumed to have the same efficiency (ie, same number of parameters in the models being compared). So perhaps this is an appropriate test when efficiency is the same between two models, and one just wants to test whether the parameter estimates differ. Furthermore, the results are statistically significant, indicating that estimates differ but it is unclear to me which is the best model.

	I believe the paper will be acceptable if the major limitation of self-reported weight is addressed by the authors, and if the use of the Hausman-type tests can be clarified.

VERSION 1 – AUTHOR RESPONSE

REVIEWER ONE COMMENTS
This article addresses a topic of importance – is there an optimal way to analyse weight change, and what are the potential implications of using different measures. I think the authors have selected their outcomes well but would suggest inclusion of another option.
Major suggestions:

1. There is a body of literature suggesting that in analyses of change in a variable, that variable should not be adjusted for at baseline (see for example Shea American Journal of Epidemiology 1998), known as the horse racing effect. However, the role of baseline weight can be important. So, I would recommend you re-run your continuous analyses also predicting follow up weight and adjusting for baseline weight. This may be relevant for the apparently discordant results for absolute versus relative continuous change for women versus men who have very different absolute weight and weight change.

OUR RESPONSE: We did not adjust for baseline weight because, as highlighted by the reviewer, it has been shown that in analyses of change variables, adjusting for the baseline value can introduce bias.1 We have now referenced the article mentioned by the reviewer. Rerunning the analyses (inappropriately) adjusting for baseline weight, neither absolute nor percentage continuous weight change is statistically significant, in men or women. We are concerned that reporting these inappropriately adjusted results is likely to confuse the reader and does not appear to add meaningfully to the research questions being addressed.

We agree with the reviewer that the discordant results for absolute versus relative change for females reflect baseline weight. We state in the results that “Notably, females had a lower mean starting weight compared to males (69.7kg compared to 83.9kg); thus for a given value of absolute weight change, the percentage change in weight was higher in females compared to males.” In our discussion we also say that “Unless baseline weight differs substantially between exposure levels, the relative and absolute weight-change measures are likely to give similar results.”

2. Your findings may also reflect the non-linear association between some predictors and weight change, such that loss is not the same as gain. Another option to consider including would be looking only at weight gain, by excluding major weight losers from the analysis. This may be an option to retain the power of a continuous analysis but not be confounded by predictors of weight loss.

OUR RESPONSE: Weight losers were not excluded from the analysis as the aim of the paper was to use methods commonly reported in papers that model weight change; often mean weight change is examined as is, without the exclusion of those losing weight.2-4 We agree with the reviewer that the findings where weight change was measured as a continuous variable may show a mixed effect as both weight loss and weight gain are captured together in the mean value. This is one of the key points in our paper; the use of mean weight change alone can obscure important directional information where high proportions of people are either losing or gaining weight within the same exposure group. This is the main justification for examining the effect of modelling weight as a categorical variable; this categorisation allows the reader to effectively exclude weight losers and hence addresses this point to some extent.

3. With your large power you could also consider a sensitivity analysis restricting the analysis to never smokers as they are a substantial confounder of the relationship which can never be fully accounted for through adjustment by simple smoking status.

OUR RESPONSE: We agree with the reviewer that smoking could be a substantial confounder of the relationship between education and weight change. As mentioned in the discussion of the paper, the aim of the study was not to measure the un-confounded relationship between education and weight change, but instead to examine the effects of different modelling of weight change measures on study results. We are concerned that adding in a sensitivity analysis restricting to never smokers is beyond the scope of the paper. However, we are happy to take editorial advice on this matter.

REVIEWER TWO COMMENTS
This paper is a well-written useful addition to the literature, as it compares different cutpoints for defining stable vs. weight change by education level. I do not know of other studies that have compared different cutpoints, although other studies have compared absolute vs. relative change.

1. Although noted in the limitations, the use of self-reported weight is a major concern, because we might expect the variance to be correlated with education level (ie, the more educated may more accurately report weight and thus weight change) and also to be non-differentially biased (for example, if the less education underreport by a greater amount than the higher-educated). I would therefore caution against over-interpretation of the results as favoring cutpoints over continuous methods. As the authors point out, we would expect a systematic under-reporting to wash out in the analyses, and while I agree, correlation of variance with education may be a greater concern, as results may be driven as much by random noise. In other words, if there is more variance in measured among the less educated than among the highly educated, then we will see greater numbers or less educated crossing upper- and lower- thresholds for defining weight change, just due to chance. This will cancel out in a linear regression, but emerge in a categorical regression.

OUR RESPONSE: We agree that a limitation of this study is that there may be misclassification in weight change that is differential with respect to the exposure (education). This may occur if (a) people with certain levels of education consistently over- or under-report weight; and/or (b) reporting of weight is generally less precise in some education groups than others, resulting in more variation in self-reporting of weight in some education groups compared to others.

In response to the first point, the correlation between self-report and measured weight change has been previously examined in the 45 and Up Study (the study population used in this paper).5 Although un-adjusted for other factors, the unpublished results from this study show that, in the 45 and Up Study population, while participants on average under-reported their weight, under-reporting of weight did not vary significantly according to education level. These unpublished results are reproduced below:

Discrepancy in Weight (kg)
(Self-reported – Measured)

Mean (95% CI) p value (Heterogeneity)

Education 0.507
No qualification -1.30 (-1.90, -0.69)
School certificate -1.13 (-1.47, -0.79)
HSC or equivalent -0.98 (-1.63, -0.32)
Trade/Certificate -1.40 (-1.79, -1.01)

Tertiary -1.52 (-1.87, -1.18)

Furthermore, misclassification of weight change groups due to systematic under-reporting of weight would only occur if the degree of under-reporting varied over time. In our study, education was measured as the highest qualification achieved (ranging from no school certificate to university degree or above) and this is likely to be stable over the course of the study period, and it is a reasonable assumption that if highly educated people under-/over-report their weight at baseline, they would most likely also do this at follow-up. Moreover, if this degree of under-reporting does not change over time, then classification according to weight change should not be affected.

With respect to the reviewer's second point, we agree that a limitation of relying on self-reported data is that if the variance in self-reporting weight varies by education then, all else being equal, relatively more people in the education group with the greater variation will cross the upper- and lower-thresholds for defining weight change; this will cancel out in a linear regression, but emerge in a categorical regression.

We have added the following text to the discussion in the paper: "Unpublished data from the weight validation study within the 45 and Up Study5 demonstrated that while people on average under-report their weight, there was no significant difference in the mean discrepancy between measured and self-reported weight according to education level. However, it is acknowledged that if precision in reporting weight change, and hence variance, varies by education level, this itself could at least partly account for the observed differences between the categorical versus continuous weight change measures. This is because the greater the variation, the higher the probability there is of crossing the upper- and lower- thresholds for defining weight change, while mean weight change remains unaffected."

2. Tables 5 and 6 seem unnecessary. We don't need to know such detail about the relationship between potential confounders and weight change. On the other hand, presenting adjusted, partially adjusted, and fully adjusted models are helpful.

OUR RESPONSE: Our secondary analysis was to examine the relationship of the various weight-change measures to the other sociodemographic and behavioural factors included in the analysis and Table 5 and 6 were included for this point. This is not the main focus of our paper and if the editorial office prefers, we can change these to be supplementary tables; our experience is that readers generally have been interested in these other variables.

3. Unless the authors can clarify the Hausman test for me, this might require someone trained in econometrics and more familiar with these tests. This might be a perfectly appropriate application of the Hausman method, I am just not certain.I am a bit confused about the use of the Hausman test here and the only reference provided for the test is the Stata users manual and I have not seen it used in epidemiologic research before.

My understanding of this test is that it tests consistency and efficiency, where a more efficient estimator is compared to a (presumed) consistent estimator. In cases where I have seen the test applied, it is used to test whether additional variance parameters need to be included in the model (for example, random effect vs. fixed effect), a robust estimator is needed (eg, due to outliers), or irrelevant alternatives in multinomial models are independent of one another (the IIA assumption). The test is really one of whether a model that is more efficient (but has unknown consistency), can substitute for a known-to-be-consistent (but less efficient) model. The main issue they are examining in this paper, however, is whether two sets of parameter estimates are statistically different, when they would be presumed to have the same efficiency (ie, same number of parameters in the models being compared). So perhaps this is an appropriate test when efficiency is the same between two models, and one just wants to test whether the parameter estimates differ.

OUR RESPONSE: Following the storing of results from the different multinomial logistic regression models, a seemingly unrelated estimation (‘suest’) post-estimation command in Stata was used followed by the ‘test’ command. This ‘test’ command can be either a Hausman test or a Wald test depending the on the specifications of the models being compared. In this case, the models being compared were multinomial logistic regression models with the same independent variables but different dependent variables and the test to compare this is the Wald test, not the Hausman test. We apologise for the confusion. This does not change the interpretation of any of the results in the paper. The text in the methods has been corrected to state “We then compared the regression coefficients across models using the different cut-points using Wald tests.” A further reference for this type of test has been added to the amended manuscript.

4. Furthermore, the results are statistically significant, indicating that estimates differ but it is unclear to me which is the best model.I believe the paper will be acceptable if the major limitation of self-reported weight is addressed by the authors, and if the use of the Hausman-type tests can be clarified.

OUR RESPONSE: The reviewer is correct; we found that the estimates differed between the multinomial logistic regression models using different cut-points for the outcome, but this does not indicate which model is best. This is a common issue with the interpretation of sensitivity analyses where the best model is not already known. We sought to bring to the reader’s attention that when choosing cut-points for categories of weight change, these cut-points may influence the size of the measure of association, however, we cannot comment on what cut-points are most appropriate. The choice of cut-points will depend on the research question of interest, what is clinically relevant and the distribution of the outcome variable, which may differ across study populations.

References:

1. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol 2005; 162(3): 267-78.
2. Ball K, Crawford D, Ireland P, Hodge A. Patterns and demographic predictors of 5-year weight change in a multi-ethnic cohort of men and women in Australia. Public Health Nutr 2003; 6(3): 269-81.
3. Brown WJ, Williams L, Ford JH, Ball K, Dobson AJ. Identifying the energy gap: magnitude and determinants of 5-year weight gain in midage women. Obes Res 2005; 13(8): 1431-41.
4. Wilsgaard T, Jacobsen BK, Arnesen E. Determining lifestyle correlates of body mass index using multilevel analyses: the Tromso Study, 1979-2001. Am J Epidemiol 2005; 162(12): 1179-88.
5. Ng SP, Korda R, Clements M, et al. Validity of self-reported height and weight and derived body mass index in middle-aged and elderly individuals in Australia. Aust N Z J Public Health 2011; 35(6): 557-63.

VERSION 2 – REVIEW

REVIEWER	Anna Peeters Baker IDI Heart and Diabetes Institute, Australia
REVIEW RETURNED	28-Apr-2014

GENERAL COMMENTS	The authors have addressed the comments well. However, they have misunderstood one of my main suggestions. "So, I would recommend you re-run your continuous analyses also predicting follow up weight and adjusting for baseline weight" Due to the "horse-racing" effect I did not suggest to adjust their

	current weight change analyses for baseline weight but to run a comparative analysis as often recommended where follow-up weight (not weight change) is the outcome variable with baseline weight as a predictor. If the results of this are basically the same this could just be mentioned as a sensitivity analysis

REVIEWER	David Shoham Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA
REVIEW RETURNED	20-May-2014

GENERAL COMMENTS	The authors have adequately addressed my major concern, which was regarding variance leading to crossing cutpoint values, and also discussed the issue of self-reported weight.